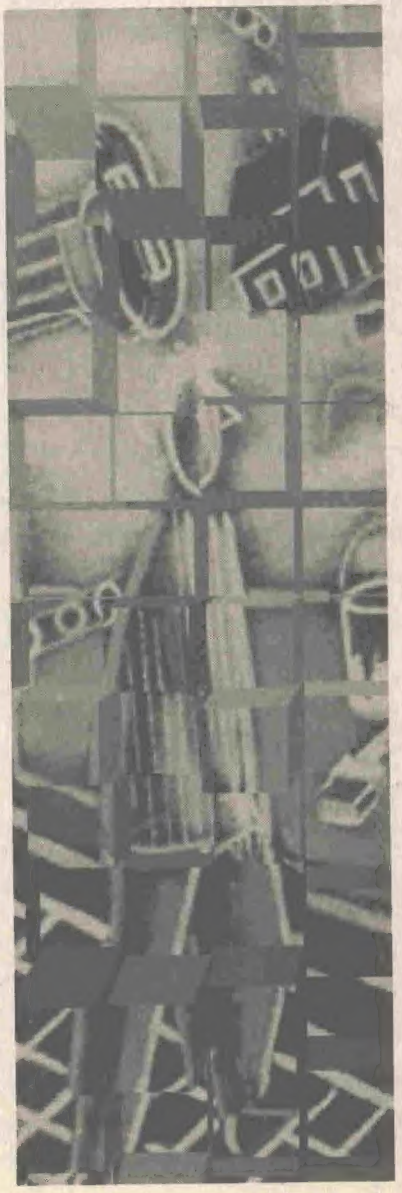
สหัส พรหมสิทธิ์

คอมพิวเตอร์วิชันกับการค้นหาภาพจากฐานข้อมูล

เป้าหมายของการค้นคว้าวิจัยในด้านคอมพิวเตอร์การหาความเป็นเอกลักษณ์ของสรรพสิ่งที่ปรากฏอยู่



เทคโนโลยีสำหรับค้นหาภาพดิจิตอลได้มีส่วนเกี่ยวข้องกับวิทยาการในด้านคอมพิวเตอร์วิชันหรือแมชีนวิชันเป็นอย่างมาก รวมทั้งเป็นสาขาหนึ่งของ AI (Artificial Intelligence) ที่ได้เริ่มมีการค้นคว้าอย่างเป็นรูปธรรมเมื่อปลายทศวรรษ 1950 เป้าหมายของการค้นคว้าวิจัยในด้านคอมพิวเตอร์วิชันก็คือการหาความเป็นเอกลักษณ์ของสรรพสิ่ง (identification of objects) ที่ปรากฏอยู่ในที่ต่าง ๆ ว่าเป็นอะไร อันเป็นการวิเคราะห์ที่รวมถึงสิ่งที่เป็นสภาพแวดล้อม (space) และช่วงเวลา (time) ที่สรรพสิ่งดังกล่าวไปปรากฏอยู่

นอกจากนี้การวิเคราะห์ในแง่ของคอมพิวเตอร์วิชันยังคลุมถึงความสัมพันธ์ (causal relationships) ของสิ่งต่าง ๆ ที่ปรากฏอยู่ ณ ที่ใดที่หนึ่งว่ามีความเกี่ยวข้องสัมพันธ์กันอย่างไร ผลงานวิจัยในด้านคอมพิวเตอร์วิชันเพื่อให้เกิดความเข้าใจในเรื่องของทัศนียภาพที่เป็นสามมิติที่ถือว่าเป็นขนานแท้ ได้เริ่มขึ้นเมื่อต้นทศวรรษ 1960 ที่ MIT อันเป็นงานวิจัยที่เป็นการวางรากฐานทั้งในเรื่องของคอมพิวเตอร์วิชันและคอมพิวเตอร์กราฟิก ดังเช่น งานวิจัยของนักวิทยาศาสตร์คอมพิวเตอร์ผู้มีนามว่า แอล. จี. โรเบิร์ต ที่ MIT เมื่อปี 1965 ซึ่งเกี่ยวกับเรื่องของ Optical and Electro-Optical Interaction Processing จากนั้นงานวิจัยด้านคอมพิวเตอร์วิชันก็ได้ก้าวเข้าสู่ยุคของ cognitive approach ในรูปของการทำความเข้าใจถึงความสัมพันธ์ของสิ่งต่าง ๆ (objects) ที่ปรากฏอยู่ในภาพโดยอาศัยความรู้ในเรื่องของสภาพแวดล้อมเข้าช่วย อันเป็นที่มาของคำว่า "knowledge-directed" vision ซึ่งก็ยังคงมีปัญหาในแง่ที่ว่าเทคนิคต่าง ๆ ที่มีใช้อยู่ยังไม่พอเพียงสำหรับการลดช่องว่างระหว่างอินพุตอิมเมจกับ symbolic descriptions

ของสิ่งที่เป็นภาพที่ต้องการ อันทำให้ต้องหาทางลดช่องว่างดังกล่าวด้วยการให้ความสนใจในเรื่องของ การรวบรวมสิ่งที่เป็น intermediate representations ซึ่งเรียกว่า **อินทรินสิคอิมเมจ** (intrinsic images) การวิจัยในเรื่องของอินทรินสิคอิมเมจเพื่ออุดช่องว่างในด้าน representation ได้เริ่มเมื่อปี 1974 โดยการนำของนักวิทยาศาสตร์คอมพิวเตอร์ผู้มีนามว่า David Marr แห่ง MIT และ Barrow กับ Tenenbaum แห่ง Stanford Research Institute จากนั้นก็ได้ก้าวเข้ามาสู่ยุคปัจจุบัน ซึ่งเริ่มต้นในราวปี 1985 ที่ได้มีการเห็นพ้องต้องกันสำหรับในวงการคอมพิวเตอร์วิชันว่าการที่จะอุดช่องว่างดังกล่าวนั้นสามารถทำได้โดยอาศัยข้อมูลที่เป็นชุดในรูปของ visual data representation ซึ่งได้รับการจัดให้อยู่ในรูปที่มีความเป็นนามธรรมมากขึ้น (a hierarchy of increasing abstraction)

คอมพิวเตอร์วิชันขั้นพื้นฐาน

ในทางปฏิบัตินั้น ระบบที่เป็นแมชีนวิชันจะต้องสามารถก่อให้เกิดสิ่งที่เป็นภาพ พร้อมกับวิเคราะห์และตีความเพื่อทำความเข้าใจภาพดังกล่าว โดยแบ่งออกเป็นขั้นต่าง ๆ ดังนี้

1. image formation
2. edge detection
3. depth perception
4. texture analysis
5. decoding color
6. object recognition
7. perspective changes

1. ในทางเทคนิคนั้น การสร้างสิ่งที่เป็นภาพ (image formation) เป็นสาขาหนึ่งของแมชีนวิชันที่ได้รับการพัฒนามาจนเรียกได้ว่าเกือบจะมีความสมบูรณ์ที่สุด ดังเช่น การบันทึกภาพด้วยกล้องถ่ายภาพโทรทัศน์ที่อาศัยแสงสะท้อนจากผิวของวัตถุที่มีรูปทรงแบบสามมิติ ข้อมูลที่ได้จากการเก็บภาพในกรณีนี้จะถูกแปลงเป็นสัญญาณไฟฟ้าที่แปรตามความเข้มของแสงสะท้อน จากนั้นก็จะใช้คอนเวอร์เตอร์เพื่อแปลงสัญญาณไฟฟ้า (อะนาล็อก) ให้เป็นข้อมูลทางดิจิตอล เพื่อใช้กับคอมพิวเตอร์ การเก็บข้อมูลที่เป็นสัญญาณไฟฟ้าเป็นการสุ่มเก็บข้อมูลเป็นระยะ ๆ ที่ตายตัว (regular intervals) เพื่อแปลงข้อมูลให้เป็นตัวเลขที่ใช้ในการแทนค่าต่าง ๆ ที่เป็นความเข้มของแสง หรือ**เกรย์สเกล** (gray scale) ตัวเลขที่เกิดจากกริดหรือตารางแบบสองมิตินั้นมีชื่อเรียกว่า เกรย์เลเวลอะเรย์ (gray level array) โดยแต่ละค่าของตัวเลขในอะเรย์ (array) จะเท่ากับหนึ่ง**พิกเซล** (pixel) หรือ picture element ของภาพที่ได้รับการดิจิไตส์แล้ว (digitized image) โดยทั่วไประบบวิชันในทาง AI จะใช้เกรย์สากลที่มีค่าตัวเลขตั้งแต่ 0 ถึง 255 โดย 0 จะแทนความมืดมากที่สุด ส่วนตัวเลข 255 จะแทนบริเวณของภาพที่สว่างที่สุด สำหรับระบบวิชันที่เป็นสีนั้น ก็จะมีการวัดปริมาณของความยาวคลื่นของแสงที่เป็นสีแดง เขียว และน้ำเงิน ซึ่งเกิดจากการสะท้อนมาจากภาพ อันเป็นการวัดแบบแยกจากกันเป็นแต่ละสี โดยมีการแปลงผลที่เกิดจากการวัดให้เป็นอะเรย์ต่าง ๆ นับเป็นจำนวนสามอะเรย์ สำหรับตัวเลขที่แสดงความสว่างที่มีค่าตั้งแต่ 0 ถึง 255 จะเห็นระบบวิชันที่เป็นสีต้องทำการประมวลผลข้อมูลมากขึ้นอีกสามเท่าในแง่ของข้อมูลต่อหนึ่งพิกเซล เมื่อเทียบกับระบบเกรย์เลเวล (gray level systems)

2. เมื่อคอมพิวเตอร์ได้รับข้อมูลซึ่งแทนความเข้มของแสงที่สะท้อนจากภาพสามมิติ มันก็จะเริ่มค้นหาความหมายของสิ่งที่แฝงอยู่ในตัวเลขข้อมูลดังกล่าว สิ่งแรกที่คอมพิวเตอร์มักจะทำก็คือ การเริ่มกรรมวิธีที่เรียกว่า edge detection เพื่อหาเส้นรอบรูปหรือลักษณะ (outlines) ของสิ่งต่าง ๆ (objects) หรือส่วนใดส่วนหนึ่งของสิ่งดังกล่าว **อัลกอริธึม**สำหรับ edge detection จะสั่งให้คอมพิวเตอร์มองหา**การเปลี่ยนแปลง**ของค่าตัวเลขที่แสดงความสว่างที่เป็น**การเปลี่ยนแปลงอย่างรวดเร็ว** ซึ่งมักจะหมายถึงการมีขอบสัน (edges) ที่เกิดจากเส้นรอบรูป (object boundaries) หรือความสูงต่ำของพื้นผิว หรือการเปลี่ยนแปลงของสี ตัวอย่างเช่น ภาพที่เป็นแบบเกรย์เลเวลนั้นตรงบริเวณที่เป็นเส้นล้อมรอบที่สำคัญ (major boundaries) จะมีการเปลี่ยนแปลงจากความมืดไปเป็นความสว่างอย่างเห็นได้ชัด เช่น ระหว่างภาพของคนที่ยืนอยู่ในที่ใดที่หนึ่งกับส่วนที่เป็นแบ็กกราวด์ **การเปลี่ยนแปลง**ดังกล่าวอาจมีปัญหาในเรื่องของความไม่ชัดเจนซึ่งเกิดจากการรบกวนที่เรียกว่า noise อันเป็นการแปรปรวนในเรื่องของความเข้มของแสงที่เกิดจากเนื้อผิว (surface texture) ของสิ่งใดสิ่งหนึ่ง หรือเกิดจากรอยตำหนิต่าง ๆ รวมทั้งที่เกิดจากการแปรปรวนทางอิเล็กทรอนิกส์ที่เกิดขึ้นในกระบวนการดิจิไตส์ (digitizing process) ซึ่งก่อนที่คอมพิวเตอร์จะสามารถทำให้สิ่งที่เรียกว่า edge detection นั้น คอมพิวเตอร์จะต้องขจัดหรือลดปริมาณของสิ่งที่เป็น noise โดยอาศัยกระบวนการที่เรียกว่า "smoothing" เพื่อให้ได้ภาพที่เรียกว่า edge picture หรือ edge map ในที่สุด กระบวนการที่เรียกว่า edge detection ที่พูดมานี้จะมีประโยชน์สำหรับงานในด้านการเข้าใจวัตถุที่เป็นสามมิติ (object recognition) แต่ก็ยังมีปัญหาว่าเทคนิคในด้าน edge detection จะให้ภาพลายเส้นแบบสองมิติ ซึ่งยังไม่มี "ความลึก" การที่จะเข้าใจสิ่งที่เรียกว่า ความลึก (depth) คอมพิวเตอร์ก็จำเป็นที่จะต้องมีข้อมูลอย่างพอเพียงในการเข้าใจสิ่งที่เป็นสามมิติ

3. ส่วนสำคัญของการมองที่ทำให้มนุษย์สามารถเห็นความลึกได้ก็คือ เทคนิคที่เรียกว่า **binocular vision** คือแต่ละข้างของตาจะส่งภาพแบบแยกจากกันไปยังสมอง ที่ทำหน้าที่แปลงต่าง ๆ ของการมองเห็นภาพที่เข้ากันไม่ได้ ให้เป็นความลึก อันเป็นวิธีที่นำไปใช้กับแมชีนวิชันได้ โดยใช้กล้องถ่ายภาพสองอันเพื่อถ่ายภาพในมุมต่าง ๆ แต่ก็จะเกิดมี**ปัญหา**ในเรื่องของ disparity ระหว่างกล้องถ่ายภาพทั้งสองในขณะเก็บภาพ อันเป็นปัญหาที่เรียกว่า correspondence problem ซึ่งเกิดจากสเตอริโอวิชันของระบบคอมพิวเตอร์ที่ไม่สามารถรับรู้ว่ามีวัตถุหรือสิ่งที่อยู่ภายนอก นอกจากตัวเลขจำนวนสองอะเรย์ที่ป้อนเข้ามาโดยมาจากกล้องถ่ายภาพแต่ละอัน คือไม่เพียงแต่

การที่จุดใดจุดหนึ่งซึ่งอยู่บนวัตถุจะไปปรากฏอยู่ในตำแหน่งต่าง ๆ ในระบบที่มีเกรย์เลเวลอะเรย์สองชุด แต่ค่าตัวเลขต่าง ๆ ดังกล่าวยังเปลี่ยนไปพร้อม ๆ กับการเปลี่ยนในด้าน perspective วิธีหนึ่งในการแก้ปัญหาที่เรียกว่า correspondence ก็คือการลดสิ่งที่เรียกว่าเกรย์เลเวลอะเรย์ซึ่งมี 2 ชุดให้เหลือแค่เพียงสิ่งที่เรียกว่า edge maps โดยอาศัยหลักของ edge detection ฯลฯ อันจะทำให้ระบบคอมพิวเตอร์สามารถหาระยะทางจากกล้องถ่ายภาพทั้งสองไปยังแต่ละจุดที่อยู่บนวัตถุ รวมทั้งการสามารถสร้างรูปทรงที่เป็นสามมิติขึ้นมาใหม่โดยสร้างทีละจุด

4. วิธีการอันหนึ่งที่ระบบแมชีนวิชันจะสามารถสร้างสิ่งที่เป็นวัตถุจากข้อมูลเป็นจำนวนมากที่เป็นตัวแทนของ external image ก็คือด้วยการอาศัยกรรมวิธีที่เรียกว่า texture analysis อันเป็นวิธีการอันเดียวกันกับที่ใช้ในเรื่อง object recognition อันเนื่องมาจากว่าแต่ละ texture จะเข้ามาปรากฏอยู่ในเกรย์เลเวลอะเรย์ของคอมพิวเตอร์ในรูปของค่าตัวเลขที่แสดงความสว่าง การเปลี่ยนแปลงอย่างกะทันหันของ texture มักจะแสดงว่ามีการเปลี่ยนแปลงในพื้นผิวทางกายภาพหรือพื้นผิวของวัตถุ เช่น การเปลี่ยนแปลงจากทุ่งหญ้าไปสู่ป่าละเมาะ อันเนื่องมาจากภาพถ่ายทางอากาศ นอกจากนี้ก็ยังมีการเปลี่ยนแปลงของ texture แบบค่อยเป็นค่อยไปที่บ่งบอกถึงการมีวัตถุที่มีรูปทรงแบบสามมิติอยู่ในภาพ

การบอกลักษณะของ texture ของพื้นผิวของสิ่งใดสิ่งหนึ่ง ทำได้สองวิธีคือ structural analysis และวิธีการทางสถิติ (statistical analysis) การวิเคราะห์ที่เรียกว่า structural analysis เป็นการมุ่งไปที่ส่วนต่าง ๆ ที่มีลักษณะเด่น รวมทั้งการดูความสัมพันธ์ของส่วนต่าง ๆ ดังกล่าว เช่น ลักษณะการเรียงเม็ดข้าวโพดเมื่อแกะเปลือกออกแล้ว รวมทั้ง texture ของเปลือกของสัปปะรด ที่มีลักษณะเฉพาะที่ไม่เหมือนผลไม้ชนิดอื่น ในการหาเอกลักษณ์ของสิ่งที่เป็นจุดเด่นนี้วิธีการทาง structural analysis จะประยุกต์ edge detector เข้ากับเกรย์เลเวลอะเรย์ เพื่อที่จะมองหาเส้นต่าง ๆ ที่ทำการแบ่งแยกส่วนต่าง ๆ ออกจากกัน (lines of demarcation) การวิเคราะห์ด้วยวิธีทางสถิติใช้สำหรับ texture ที่ไม่มีลักษณะเด่น หรือใช้กับภาพที่ไม่สามารถแลเห็นรายละเอียดต่าง ๆ เช่นภาพถ่ายจากดาวเทียม หลักเกณฑ์การทำงานของวิธีทางสถิติก็คือการหาความสัมพันธ์ระหว่างจุดใดจุดหนึ่งที่เป็นพิกเซล (pixel) กับสิ่งที่อยู่ใกล้เคียงกับจุดดังกล่าวซึ่งอยู่ในเกรย์เลเวลอะเรย์ โดยการหา probability ที่ว่า ค่าความเข้มของแสงของพิกเซลดังกล่าวจะเหมือนกับจุดที่อยู่ข้างเคียง อันเป็นการเปรียบเทียบคุณสมบัติของพื้นผิวหรือของ texture ของสิ่งใดสิ่งหนึ่งที่ดูจากทิศทางการเรียงตัว (directional quality) ของสิ่งที่เป็นส่วนประกอบของ texture ดังเช่นกรณีของมะพร้าวที่ปอกเปลือกแล้วและเหลือแต่ขนหุ้มอยู่ภายนอก ซึ่งสามารถนำกรรมวิธีทางสถิติมาหาค่าความเข้มของแต่ละพิกเซลเมื่อเทียบกับจุดที่อยู่ใกล้เคียงโดยดูจากในแนวขนานและในแนวตั้งฉาก อันเป็นการหา directional quality ของผลมะพร้าวที่ไม่มีเปลือกหุ้มอยู่ คุณสมบัติอื่นในด้าน texture ที่นำสถิติมาวิเคราะห์ได้ก็มีอาทิเช่น ความหยาบ (roughness) ความสม่ำเสมอของผิวเนื้อ (regularity) และความแตกต่าง (contrast)

5. ปัญหาอันเป็นความยากอีกเรื่องในระบบแมชีนวิชันก็คือ การหาเอกลักษณ์ของสิ่งใดสิ่งหนึ่งโดยดูจากสี ซึ่งเป็นคุณสมบัติประจำตัวของสิ่งนั้น อย่างไรก็ตาม สีที่น่าจะเป็นคุณสมบัติประจำตัวดังกล่าวก็ยังอยู่ภายใต้อิทธิพลในเรื่องของแสงและเงา (illumination) ที่เข้ามาทำให้การหาเอกลักษณ์ของสิ่งต่าง ๆ ยากขึ้น อันทำให้จำเป็นต้องมองคุณสมบัติของสีโดยละเอียดมากขึ้นในแง่ของ hue ซึ่งหมายถึงชนิดของสี เช่น แดง ส้ม น้ำเงิน ฯลฯ รวมทั้งความเข้ม (intensity) อันเป็นการวัดความสว่างและ saturation ซึ่งหมายถึงความบริสุทธิ์ของสี ในแง่ของอิทธิพลนั้น illumination จะเข้ามาเปลี่ยนแปลงความเข้มและ saturation ของสีที่มองเห็น แต่สิ่งที่เรียกว่า hue จะไม่อยู่ภายใต้อิทธิพลของการเปลี่ยนแปลงของแสง กล้องถ่ายภาพในระบบแมชีนวิชันสามารถบันทึกได้เฉพาะส่วนของสีที่เป็น แดง เขียว และน้ำเงิน (RGB) เมื่อเป็นเช่นนี้ การที่จะหาเอกลักษณ์ของสิ่งที่เรียกว่า hue ของวัตถุก็จำเป็นต้องแปลงค่าตัวเลขของสีทั้งสามให้เป็นการวัดในแง่ของชนิดของสี (hue) ความเข้ม และ saturation ซึ่งต้องอาศัยกรรมวิธีทางคณิตศาสตร์หรืออาศัยไดอะแกรมที่เรียกว่า **three dimensional model of colors** สำหรับในระบบแมชีนวิชัน

6. การที่จะหาเอกลักษณ์ของสิ่งใดสิ่งหนึ่ง (เมื่อคอมพิวเตอร์ได้รู้คุณสมบัติของสิ่งนั้นอยู่บ้างแล้ว) สิ่งที่ทำหน้าที่เป็นแมชีนหรือคอมพิวเตอร์ก็จะต้องจับคู่หรือแมตช์ (match) สิ่งที่เป็นรูปร่างเอาต์ไลน์ (outline) สี หรือ texture ที่ได้มาจากภายนอก ให้เข้ากับโมเดลที่เก็บไว้ในเมมโมรี ซึ่งมีอยู่สามวิธี คือ

6.1 เท็มเพลตแมตชิง (template matching)

6.2 ฟีเจอร์เอ็กซ์แทรคชัน (feature extraction)

6.3 รีเลชันแนล (relational approach)

6.1 กรรมวิธีหาเอกลักษณ์ที่ง่ายที่สุดก็คือ แบบที่เรียกว่าเท็มเพลตแมตชิง ในกรณีนี้คอมพิวเตอร์จะทำการเลือกคู่ที่ดีที่สุด (best match) สำหรับเอาต์ไลน์ของสิ่งใดสิ่งหนึ่ง โดยดูจากเอาต์ไลน์ต่าง ๆ ที่เก็บไว้ในเมมโมรี (stored outlines) หรือที่เรียกว่า เท็มเพลต ซึ่งเก็บสิ่งต่าง ๆ ไว้ในจำนวนจำกัด อย่างเช่นในกรณีของ optical character readers ที่สามารถอ่านและเข้าใจตัวพิมพ์ต่าง ๆ ที่เก็บข้อมูลไว้ในเท็มเพลตเดียว หรือโรบอตที่ใช้ในอุตสาหกรรมที่สามารถเลือกหยิบชิ้นส่วนต่าง ๆ จาก conveyor belt ซึ่งในกรณีนี้จะมีการใช้หลาย ๆ เท็มเพลตเพื่อเก็บข้อมูลของสิ่งที่เป็นชิ้นส่วนไว้มากมายหลายแบบ ในทั้งสองกรณีดังกล่าวตำแหน่งของการวางสิ่งที่จะให้กล้องถ่ายภาพอ่านข้อมูลในขณะใช้งานจะมีการจำกัดในแง่ของ perspectives ซึ่งถ้าหากผิดไปจากที่กำหนดไว้ (เช่นมีการหมุนเอียง) ก็จะทำให้การปรากฏตัวของสิ่งที่ต้องการจะให้คอมพิวเตอร์อ่านหรือมองแตกต่างจากที่ปรากฏในเท็มเพลต จนทำให้หาเอกลักษณ์ไม่ได้ว่าอะไรเป็นอะไร

6.2 ฟีเจอร์เอ็กซ์แทรคชัน ซึ่งมีความยืดหยุ่นกว่าแบบเท็มเพลตแมตชิง เพราะเป็นระบบที่จำแนกสิ่งต่าง ๆ โดยอาศัยการสอบวัดหรืออาศัยคุณลักษณะบางอย่างของสิ่งที่ต้องการจะหาเอกลักษณ์ โดยทั่วไปแล้วระบบที่เป็นฟีเจอร์เอ็กซ์แทรคชันจะต้องการข้อมูลอย่างน้อยสองฟีเจอร์ เช่น สีและรูปร่าง เพื่อใช้ในการบอกข้อแตกต่างระหว่างสตรอเบอรี่กับถั่วชนิดหนึ่ง (string beans) และกล้วยหอม แต่ถ้าจะให้บอกข้อแตกต่างระหว่างสตรอเบอรี่กับมะเขือเทศ ก็จะต้องบอกฟีเจอร์เพิ่มเข้าไปอีกอันเช่น texture

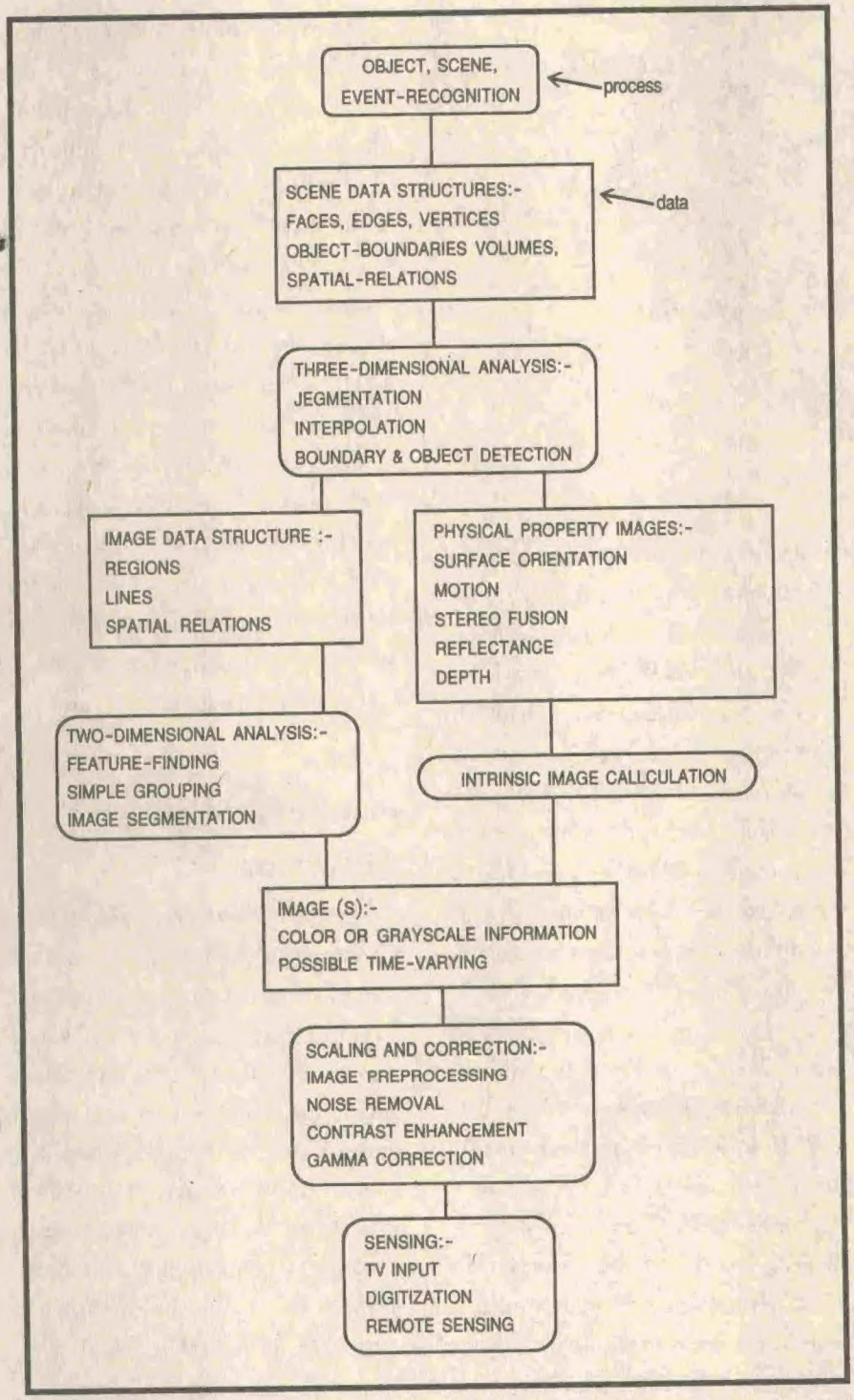
6.3 วิธีที่สามซึ่งเป็นแบบ relational approach นั้นสามารถทำให้คอมพิวเตอร์สิ่งที่มีความซับซ้อนมาก ๆ ได้โดยที่สิ่งต่าง ๆ ดังกล่าวอาจถูกป้อนเข้ามาในคอมพิวเตอร์ด้วยมุมมองที่ต่าง ๆ กัน แมชีนวิชันในลักษณะนี้จึงเป็นการมองความสัมพันธ์ของส่วนต่าง ๆ ที่ประกอบกันขึ้นมาเป็นสิ่งใดสิ่งหนึ่ง แทนที่จะมองเฉพาะแค่รายละเอียดของส่วนต่าง ๆ (descriptions of the parts) เช่นในกรณีของร่างกายคน ซึ่งอาจสูงหรือเตี้ย เป็นชายหรือหญิง นอนหรือนั่งหรืออยู่ในอิริยาบทอื่น ซึ่งสำหรับในกรณีของการใช้วิชันแบบ relational approach นั้น ภายในคอมพิวเตอร์จะมีโมเดลที่เรียกว่า คอนเน็กติวิตี้แม็พ (connectivity map) ที่แสดงส่วนต่าง ๆ ของร่างกายคนซึ่งมีความเชื่อมโยงสัมพันธ์กันในรูปต่าง ๆ เช่น "is connected to", "consists of" หรือ "is a part of" กรรมวิธีในการเข้าใจสิ่งที่เป็นวัตถุแบบสามมิติที่เรียกว่า relational approach นี้จึงมีที่ใช้ทั้งในด้านแมชีนวิชันและการค้นหาภาพที่เป็นดิจิตอล (searching for digital pictures)

7. ในกรณีที่ต้องการจะให้แมชีนวิชันเกิดประโยชน์อย่างแท้จริง ผู้วิจัยในด้าน AI ต้องเรียนรู้ที่จะเก็บคอนเซ็ปต์ต่าง ๆ ที่สะท้อนความหลากหลายไว้ในคอมพิวเตอร์ อันเนื่องจากว่าความหลากหลาย (diversity) ดังกล่าวได้เกิดมีอยู่ทั่วไปในโลกของความเป็นจริง หนึ่งในเรื่องของความหลากหลายซึ่งถือว่าเป็นของยากที่จะให้คอมพิวเตอร์จะรับรู้ก็คือ **perspective changes** ในแง่ที่ว่า**การปรากฏตัวของสิ่งใดสิ่งหนึ่งนั้นอาจเป็นไปได้หลายรูปแบบ** เช่น ผลส้มที่ปอกเปลือกแล้วกับที่ยังไม่ได้ปอกเปลือก ซึ่งคอมพิวเตอร์ยังมีความรู้ในด้านนี้น้อยมาก นอกจากนี้ก็ยังมีปัญหาในเรื่องการซ้อนทับกันของชิ้นส่วนสองชั้นใน conveyor belt ในโรงงานที่คอมพิวเตอร์อาจเข้าใจว่าเป็นชิ้นส่วนเดียวกัน แต่ไม่รู้ว่าเป็นอะไร รวมทั้งปัญหาในเรื่องของการเปลี่ยนแปลงในด้าน illumination ที่สามารถสร้างความสับสนให้กับการทำงานของระบบแมชีนวิชัน

รากฐานของแมชีนวิชันหรือคอมพิวเตอร์วิชันที่พูดมาในข้างต้นนี้ ได้มีอยู่หลายส่วนที่ใช้ได้กับ "**การค้นหาภาพที่เป็นดิจิตอล**" ซึ่งจำเป็นต้องกล่าวอย่างรวมกันไปในขั้นแรก เพราะแยกออกจากกันได้ยาก ดังจะเห็นได้จากประวัติของนักวิจัยในเรื่องของการค้นหาภาพที่เป็นดิจิตอล (ซึ่งเคยปรากฏในวารสารไมโครคอมพิวเตอร์) เช่นที่ปรากฏในโครงการของมหาวิทยาลัยแห่งแคลิฟอร์เนียที่เบิร์กเลย์ ที่เรียกว่า Digital Library Project นั้นส่วนใหญ่จะเป็นผู้เชี่ยวชาญในด้านคอมพิวเตอร์วิชัน ซึ่งถือว่าเป็นแขนงหนึ่งของวิทยาศาสตร์คอมพิวเตอร์ที่เบิร์กเลย์ รายชื่อของนักวิจัยในกรณีของเบิร์กเลย์ก็มีอาทิเช่น David Forsyth, Jitendra Malik และ Robert Wilensky สำหรับในช่วงก่อนถึงเดือนมิถุนายน ปี 1997

เทคโนโลยีในการค้นหาภาพจากฐานข้อมูล

การค้นหาข้อมูลเช่นที่เป็นภาพดิจิตอลเพื่อสนองความต้องการของผู้ใช้โดยทั่วไป ไม่ว่าจะเป็นการค้นหาจากในเน็ตเวิร์กเช่นอินเทอร์เน็ต หรือจากในฐานข้อมูลในโครงการ Digital Library Project ที่เบิร์กเลย์ (http://elib.cs.berkeley.edu) ได้มีจุดมุ่งหมายในอันที่จะให้การค้นหาภาพต่าง ๆ ดังกล่าวเป็นไปอย่างอัตโนมัติในรูปของ fully automated image analysis เช่นด้วยการอาศัยคอมพิวเตอร์โปรแกรมที่สามารถบอกว่าภาพที่เก็บไว้ในรูปของดิจิตอลเป็นจำนวนมากนั้น ได้มีภาพที่บรรจุสิ่งที่ผู้ใช้ข้อมูลรายใดรายหนึ่งต้องการหรือไม่ โดยโปรแกรมดังกล่าวจะต้องสามารถหาสิ่งที่ผู้ใช้ต้องการได้อย่างถูกต้องและ**ไม่ขึ้นกับ**เรื่องของสีสัน ขนาด การปรากฏตัวของสิ่งที่ต้องการจะค้นหา หรือมุมมองที่มุ่งไปยังสิ่งที่ต้องการจะค้นหา อย่างไรก็ตาม การค้นหาเพื่อหาสิ่งต่าง ๆ ที่ปรากฏอยู่ในภาพในปัจจุบันยังรุดหน้าไปไม่ไกล สิ่งที่นักวิจัยส่วนใหญ่พอจะทำได้ก็คือ การแก้ปัญหาในเรื่องนี้ในทิศทางต่าง ๆ กัน รวมทั้งการอาศัยกรรมวิธีแบบผสมในรูปของคอมพิวเตอร์วิชันกับสิ่งที่เรียกว่า automated reasoning เพื่อเพิ่มขีดความสามารถในการวิเคราะห์หาฟีเจอร์ต่าง ๆ ที่ปรากฏอยู่ในภาพ

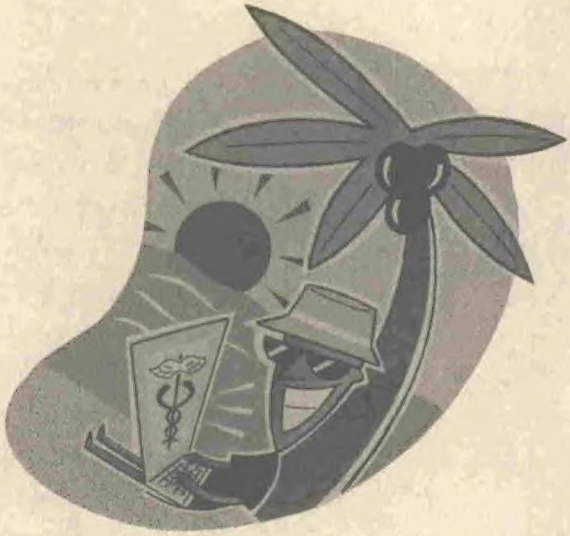


รูปที่ 1 แสดง processes (⬭) และ data representations (▭) สำหรับระบบคอมพิวเตอร์วิชันโดยทั่วไป process สามารถเริ่มจาก input data และมาที่ symbolic description (data driven scheme) หรือในทางตรงข้าม (expectations guide processing) สำหรับในแง่ของการควบคุม (control) ทำได้ทั้งสองทาง โดยในแต่ละขั้นของ processing จะรวมถึงความสัมพันธ์ระหว่าง image information กับโลกภายนอก ในกรณีของสิ่งที่มีชีวิต (natural systems) สมมติฐานข้างต้นอาจมีมาตั้งแต่เกิด หรือเกิดจากการเรียนรู้ในภายหลัง

คอมพิวเตอร์โปรแกรมยุคใหม่สามารถวิเคราะห์เนื้อหาของภาพที่จะเรียกออกมาจากฐานข้อมูลได้หลายวิธี กล่าวคือ บางโปรแกรมก็จะมุ่งไปที่ภาพซึ่งสามารถแมตช์ (match) ได้อย่างใกล้เคียงกับภาพที่เป็นตัวอย่าง นอกจากนี้ก็ยังมีโปรแกรมที่มุ่งไปที่ภาพปรากฏโดยทั่วไป (general appearance) รวมทั้งโปรแกรมที่มองหาสิ่งใดสิ่งหนึ่งที่มีเอกลักษณ์ตามที่กำหนดไว้ เช่น ห้องน้ำ หรือสิงสาราสัตว์ รากฐานของการค้นหาสิ่งที่เป็นภาพ (image retrieval) เท่าที่พูดมานี้พอสรุปได้ว่า มีอยู่สามแนวทาง อันแรกก็คือการหาภาพที่ลักษณะเหมือนกับตัวอย่างที่กำหนดให้, วิเคราะห์ภาพในรูปของการดูรายละเอียดเช่นของ texture รวมทั้งใช้วิธีการค้นหาสิ่งต่าง ๆ เหมือนกับที่มนุษย์ใช้ อันได้แก่ แพตเทิร์นแมตชิง รวมทั้งวิธีการอื่น ๆ ซึ่งจะอธิบายในขั้นรายละเอียด ดังนี้

แพตเทิร์นแมตชิง

หลักการทำงานของแพตเทิร์นแมตชิงเป็นการเปรียบเทียบอินพุตเซตของบิต (ซึ่งอาจเป็นภาพ เป็นขบวนของคาแร็กเตอร์หรือเป็นข้อมูลอื่น) กับเซตของเท็มเพลตที่มีความสัมพันธ์กับแพตเทิร์นต่าง ๆ ซึ่งต้องการจะค้นหา การค้นหาดังกล่าวจะประสบความสำเร็จก็ต่อเมื่อจำนวนของอินพุตบิตที่แมตช์กับบิต (bits) ในเท็มเพลตมีสูงเกินกว่าขีดจำกัดที่กำหนดไว้ ที่ใช้ของแพตเทิร์นแมตชิง มีอาทิเช่น การอ่านเพื่อทำความเข้าใจข้อความที่เป็นลายมือ (handwriting recognition) การหาเอกลักษณ์ของใบหน้า (face identification) การค้นหาข้อมูลจากฐานข้อมูล และการค้นหาสิ่งที่เป็นเป้า (เช่น เป้าหมายทางทหาร) โดยอัตโนมัติ สำหรับในการค้นหาภาพของสิ่งที่เป็นเป้าหมาย (target recognition) นั้น การท้าทายสูงสุดของการค้นคว้าในด้านนี้ก็คือการสามารถเปรียบเทียบภาพที่เป็นอินพุตกับเท็มเพลตที่มีจำนวนเป็นพันได้อย่างรวดเร็ว เท็มเพลตอันใดอันหนึ่งอาจทำหน้าที่ในการแสดงภาพด้านหน้าหรือด้านข้างของรถแบบใดแบบหนึ่ง ภายในแต่ละภาพ (image) จะมีจำนวนพิกเซล (pixels) นับ



เป็นพัน โดยสิ่งที่เป็นเป้าหมายจะอยู่ในลักษณะใดก็ได้ ความต้องการในการหาเป้าหมายให้ได้อย่างรวดเร็วสำหรับการใช้งานในทางทหารทำให้ระบบคอมพิวเตอร์ต้องสามารถเปรียบเทียบสิ่งต่าง ๆ ให้ได้ในอัตราหลาย ๆ trillion โอเปอเรชันต่อวินาที (1 trillion = 10^{12}) อันเนื่องจากว่าจะต้องมีการเปรียบเทียบพิกเซลทั้งหมดของภาพที่เป็นอินพุตกับเท็มเพลตเป็นจำนวนมาก ตัวอย่างหนึ่งของงานวิจัยเพื่อใช้คอมพิวเตอร์ในการค้นหาเป้าหมายเช่นที่ได้พูดมานี้ก็คือผลงานวิจัยของแผนกวิศวกรรมไฟฟ้าของ UCLA ซึ่งนำโดยนักวิศวกรสองคนคือ John Villasenor กับ William Mangione Smith ซึ่งได้รับการสนับสนุนจาก DARPA (Defense Advanced Research Projects Agency ของสหรัฐฯ สิ่งที่เป็นรูปธรรมอันเกิดจากการวิจัยนี้ก็คือการสร้างต้นแบบของระบบเรคกอกนิชัน (recognition system) โดยใช้ฮาร์ดแวร์แบบปรับตัวเองได้ (configurable hardware) ซึ่งทำให้สามารถประหยัดการลงทุนในด้านฮาร์ดแวร์ได้มากในแง่ที่ว่า ฮาร์ดแวร์แบบใหม่นี้สามารถที่จะปรับตัวของมันให้เข้ากับรายละเอียดของแต่ละเท็มเพลตอย่างอัตโนมัติ คือในกรณีที่พิกเซลต่าง ๆ ซึ่งอยู่ในเท็มเพลตอันใดอันหนึ่ง ไม่มีผลกับการแมตช์ที่จะนำไปสู่คำตอบในขั้นสุดท้าย ก็จะถูกละเว้นจากการประมวลผลข้อมูลโดยปริยาย อันเป็นความสามารถพิเศษที่ไม่มีในฮาร์ดแวร์แบบธรรมดาที่ใช้กันอยู่ทั่วไป เพราะพิกเซลต่าง ๆ ซึ่งจะต้องมีการละเว้นไม่ประมวลผลนั้นจะแตกต่างกันไปในแต่ละเท็มเพลต อันเป็นสิ่งที่ฮาร์ดแวร์แบบอื่น ๆ ทำไม่ได้ การหาทางใช้ประโยชน์จากความยืดหยุ่นของ **configurable hardware** ยังทำได้อีกในแง่ของการจูนฮาร์ดแวร์แบบนี้ให้อาศัยความได้เปรียบในแง่ของความคล้ายคลึงกันของเท็มเพลตต่าง ๆ เพื่อเพิ่มประสิทธิภาพในการประมวลผลข้อมูล อันเนื่องจากว่า ฮาร์ดแวร์ชนิดนี้สามารถประมวลผลให้แก่เท็มเพลตที่เป็นชุดได้ในรูปที่ขนานกันโดยอาศัย comparison unit เพียงหนึ่งตัวต่อหนึ่งพิกเซลที่มีค่าเหมือนกัน สำหรับในเท็มเพลตต่าง ๆ ที่อยู่ในชุด เช่น แทนที่จะต้องใช้วงจรฮาร์ดแวร์ธรรมดาถึง 8 ชุดในการพิจารณาพิกเซลเพียงจุดเดียวจากเท็มเพลตต่าง ๆ ที่มีอยู่ 8 ซึ่งสำหรับในกรณีที่ใช้วงจรฮาร์ดแวร์ที่เป็นแบบ configurable เช่น FPGA (Field-Programmable Gate Arrays ที่ได้เคยพูดรายละเอียดไว้ในวารสารนี้) ก็จะใช้เพียงวงจรเดียวสำหรับในการพิจารณาพิกเซลในเท็มเพลตอันหนึ่งแล้วส่งผลที่ได้ไปยังเท็มเพลตอื่น ๆ อีกเจ็ดอัน อย่างไรก็ตาม เทคโนโลยีในการค้นหาภาพจากฐานข้อมูลเท่าที่มีใช้อยู่ในปัจจุบัน อันไม่ว่าจะเป็นระบบ QBIC (Query by Image Content) ของ IBM หรือ Photobook ของ MIT รวมทั้งกรรมวิธีที่เรียกว่า แพตเทิร์นแมตชิง ต่างก็ยังย่ำเท้าอยู่กับภาพสองมิติ แม้ว่าจะเป็นระบบค้นหาข้อมูลที่เป็นภาพซึ่งได้รับการยอมรับในแง่ของการใช้งานว่ามีประสิทธิภาพอันสูงก็ตาม อันทำให้ต้องมีการค้นคว้ากันต่อไปอีก เพื่อสนองความต้องการผู้ใช้ในการค้นหาข้อมูลที่เป็นภาพ ซึ่งประกอบด้วยสิ่งต่าง ๆ ที่เป็นสามมิติเป็นส่วนใหญ่ การวางรากฐานสำหรับ content based retrieval จึงต้องดำเนินต่อไปในสถาบันต่าง ๆ ทั้งที่อยู่ในสหรัฐฯ และในประเทศอื่น แนวทางในการวิจัยในด้าน content base retrieval เช่นในสหรัฐฯ ที่ผู้เขียนเห็นว่าน่าสนใจมากและจะนำมากล่าวเป็นรายละเอียดเพิ่มเติมในที่นี้ก็คือ ผลการวิจัยของทีมงานที่เบิร์กเลย์ ซึ่งประกอบด้วย David Forsyth, Jitendra Malik และ Robert Wilensky ที่มุ่งไปในการสร้างวิธีการในการค้นหาภาพสามมิติจากฐานข้อมูลใน Digital Library Project อันเป็นโครงการที่สานต่อจากที่ได้มีผู้บุกเบิกไว้บ้างแล้วในอดีตโดยนักวิทยาศาสตร์คอมพิวเตอร์ผู้มีนามว่า Ginger Ogle กับ Chad Carson

content-based retrieval ของเบิร์กเลย์

งานวิจัยเพื่อค้นหาภาพดิจิตอลจากฐานข้อมูลของมหาวิทยาลัยแห่งแคลิฟอร์เนียที่เบิร์กเลย์ได้เริ่มมานานพอสมควรก่อนที่จะเกิดมี Digital Library Project (DLP) ซึ่งมี Robert Wilensky เป็นหัวหน้าทีมในปัจจุบัน ระบบเก็บข้อมูลที่เป็นภาพในระยะแรก ๆ ซึ่งสร้างขึ้นมาโดยนักวิจัยสองท่าน (Ginger Ogle และ Chad Carson) ได้มีภาพถ่ายต่าง ๆ เก็บไว้กว่า 600,000 ภาพ โดยมีอยู่ราว 50,000 ภาพที่มาจากแหล่งต่าง ๆ รวมทั้งภาพถ่ายบนพื้นดินและทางอากาศ ซึ่งในปัจจุบันได้มีให้ผู้สนใจสามารถใช้ภาพดังกล่าวได้ในรูปของออนไลน์ โดยสามารถเรียกดูจากเวบไซท์ของโครงการนี้ อินเตอร์เฟสสำหรับผู้ใช้ในการค้นหาภาพในกรณีของ DLP นี้ได้เปิดโอกาสให้ผู้ค้นหาข้อมูลสามารถตั้งคำถามต่าง ๆ ในอันที่จะเรียกข้อมูลที่เป็นสิ่งต่าง ๆ (things) รวมทั้งที่อยู่ในรูปขององค์ประกอบต่าง ๆ หรือสตัฟ (stuff) ออกมาดู ที่สำคัญก็คือว่าการเรียกดูข้อมูลที่เป็น

สิ่งต่าง ๆ นั้นทำได้ยาก จึงได้มีการหาทางออกด้วยการเปิดโอกาสให้ผู้ใช้เรียกหาภาพซึ่งมีสิ่งที่เป็นองค์ประกอบต่าง ๆ ซึ่งสามารถนำองค์ประกอบดังกล่าวมารวมตัวกันจนเกิดเป็นสิ่งต่าง ๆ ที่ผู้ใช้ต้องการ ความเป็นมาของนักวิจัยที่เบิร์กเลย์สำหรับในเรื่องนี้ได้มีว่าการแสดงออก (หรือการตั้งคำถาม) สำหรับสิ่งต่าง ๆ ควรเป็นไปในรูปของการนำองค์ประกอบต่าง ๆ (stuff) มาประกอบกันขึ้นให้เกิดเป็นตัวเป็นตนของสิ่งดังกล่าว คำถามเพื่อค้นหาสตัฟที่มีประโยชน์ที่สุดจึงเป็นคำถามที่สามารถใช้ในการสร้างตัวแทน (representations) ของสิ่งต่าง ๆ อินเตอร์เฟสของผู้ใช้ในกรณีนี้จึงควรมีออปชันต่าง ๆ สำหรับให้ผู้ใช้เลือกใช้ในแง่ที่ว่าผู้ใช้อาจถามหาสิ่งต่าง ๆ หรือมีปัญหาเพราะการเรียกหาสิ่งดังกล่าวไม่เป็นที่น่าพอใจผู้ใช้ก็สามารถที่จะกำหนดวิธีการในการรวมตัวของสตัฟ (spatial assemblies of stuff) ที่เป็นจะทางนำไปสู่สิ่ง (things) ที่ผู้ใช้ต้องการได้ในที่สุด อันเป็นกรรมวิธีในการค้นหาสิ่งที่เป็นรูปสามมิติหรือภาพที่เหมือนกับข้อย่อยที่ (3) ของข้อ 6. ที่ว่าด้วย relational approach ซึ่งได้พูดมาแล้วในตอนต้น

ในการทดสอบความไว้วางใจได้ของระบบหาข้อมูลที่เป็นภาพของ DLP นั้น หนึ่งในทีมงานวิจัยที่เบิร์กเลย์ได้ทดลองตั้งคำถามเพื่อหาภาพ windsurfing จากฐานข้อมูล โดยให้คอมพิวเตอร์หาภาพที่มีสีน้ำเงินแกมเขียวและร้อยละ 30 ของส่วนผสมของสีดังกล่าวต้องเป็นสีของท้องฟ้าหรือทะเล รวมทั้งมีจุดสีเหลืองอย่างน้อยหนึ่งจุดติดอยู่บนแถบสีเหลืองของใบเรือ ผลที่ได้ออกมาก็คือว่าภาพส่วนใหญ่ที่ได้จากฐานข้อมูลสอดคล้องกับคำถามที่ป้อนเข้าไป รวมทั้งการมีภาพที่ไม่ตรงกับที่ต้องการหลุดออกมาให้เห็น และแม้กระทั่งภาพที่เกี่ยวกับ windsurfing อื่น ๆ ที่มีอยู่ในฐานข้อมูลก็ไม่ได้นำออกมาให้ผู้ใช้แลเห็น การค้นหาข้อมูลเช่นในกรณีนี้จึงเป็นการประนีประนอมระหว่างความแม่นยำกับความสามารถในการระลึก (recall) ซึ่งหมายถึงว่า การค้นหาข้อมูลเพื่อให้ได้เฉพาะรายการที่ต้องการจริง ๆ นั้น ก็จะมีบางรายการที่หายไป และถ้าจะต้องค้นหาข้อมูลทุกรายการตามที่ต้องการ ก็จะมีสิ่งที่ไม่ต้องการติดออกมาด้วย ตัวอย่างเช่นที่พูดมานี้ได้แสดงให้เห็นว่า กลไกในการค้นหาข้อมูลที่อยู่ในรูปของสตัฟที่มีอานุภาพจะสามารถให้ข้อมูลอย่างกว้าง ๆ ในขั้นแรกที่ถือว่าใช้การได้เพื่อเป็นทางไปสู่เป้าหมายของการหาภาพที่มีสิ่งต่าง ๆ บรรจุอยู่ในนั้น จึงเป็นดุลยพินิจของผู้ใช้ว่าเมื่อใดจึงควรเปลี่ยนจากการใช้คำสั่งเพื่อให้คอมพิวเตอร์ค้นหาสิ่งใดสิ่งหนึ่งไปเป็นคำถามเพื่อให้คอมพิวเตอร์หาสิ่งที่เป็นองค์ประกอบ (หรือสตัฟ) ออกมาให้ ระบบ content-based retrieval ของเบิร์กเลย์ในปัจจุบันได้เปิดโอกาสให้ผู้ใช้ค้นหาข้อมูลของสิ่งต่าง ๆ จากฐานข้อมูลได้เพียงไม่กี่อย่าง เช่น ม้า เรือใบ และคนที่ไม่มีเครื่องนุ่งห่ม สำหรับในการกำหนดสิ่งที่เป็นสตัฟนั้นผู้ใช้ต้องบอกร้อยละของสีต่าง ๆ ที่คาดว่าจะมีอยู่ในภาพที่จะเรียกออกมาดู พร้อมด้วยการบอกจำนวนและขนาดของ "color-blobs" ที่คิดว่าจะมี (คำว่า blob จะหมายถึงคุณลักษณะของสิ่งต่าง ๆ ที่แสดงท่วงท่า (position) รูปร่าง สี และ texture ส่วนคำบอกเล่าอื่น ๆ ที่ผู้ใช้ฐานข้อมูลต้องแจ้งให้คอมพิวเตอร์รู้ก็ยังมีอาทิ เช่น สิ่งที่เป็นภาพขอบฟ้า (horizon) หรือไม่ ขั้นสุดท้ายก็คือการที่ผู้ใช้สามารถค้นหาข้อมูลที่เรียกว่า metadata อันเป็นรายละเอียดเพิ่มเติมที่เก็บไว้ในฐานข้อมูลโดยติดอยู่กับภาพ เนื้อหาของรายละเอียดดังกล่าวจะประกอบด้วย

1. ข้อความประกอบภาพสั้น ๆ
2. ผู้ถ่ายภาพ
3. วัน, เดือน, ปีที่ถ่ายภาพ ฯลฯ

ปัญหาที่เกิดขึ้นในการทำวิจัยในการค้นหาข้อมูลที่เป็นภาพดิจิตอลตามที่ทีมงานที่เบิร์กเลย์ได้ประสบอยู่นั้น จะเหมือนกับที่ได้กล่าวไว้แล้วในตอนต้น ในหัวข้อที่เรียกว่า **"คอมพิวเตอร์วิชันขั้นพื้นฐาน"** จึงไม่ขอกล่าวซ้ำ แต่จะขอกล่าวต่อไปในเรื่องของ perceptual grouping ซึ่งหมายถึงความสามารถของคอมพิวเตอร์โปรแกรมในการรวมสิ่งที่เป็นสตัฟ (stuff) เข้าด้วยกันเป็นกลุ่ม เพื่อให้เกิดเป็นสิ่งต่าง ๆ ในภาพ ปัญหาในเรื่องของ perceptual grouping นี้ได้มีการศึกษากันมากในแง่ของวิชันของมนุษย์ โดยนักวิจัยที่อยู่ในกลุ่มของ Gestalt school of psychology อันเป็นสถาบันทางความคิดที่ได้ให้ข้อคิดหลาย ๆ อย่างที่สามารถนำมาใช้ในการสร้างเงื่อนไขว่าเมื่อใดส่วนต่าง ๆ ของภาพจึงจะมีโอกาสมากที่สุดที่จะเกิดจากสิ่งเพียงสิ่งเดียวที่อยู่ในทัศนียภาพ (scene) ความคล้ายคลึงของสีหรือ texture จึงมีบทบาทมากสำหรับในเรื่องนี้ ดังจะเห็นได้จากกลุ่มของชนเผ่าต่าง ๆ ที่เป็นมนุษย์ที่สามารถแสดงออกได้ด้วยสีของผิวของแต่ละเผ่าพันธุ์ หรือ texture ที่ใช้ในการแสดงคุณสมบัติภายนอก เช่น ผม และส่วนประกอบบนใบหน้า (หู ตา จมูก ฯลฯ) ตลอดจนลักษณะของผิวพรรณของร่างกาย ในอีกระดับหนึ่งของการจัดเป็นกลุ่มก็มีอาทิเช่น การสร้างสิ่งที่เป็นสามมิติจากภาพสองมิติที่มีสภาพเป็น symmetric รอบแกนใดแกนหนึ่ง เช่นกรณีของแจกัน นักวิจัยในด้านคอมพิวเตอร์วิชันของเบิร์กเลย์ได้ใช้เวลาอยู่หลายปีในอันที่จะแปลงกฎเกณฑ์ต่าง ๆ ที่เกิดจากสถาบันที่เรียกว่า Gestalt ให้เป็นอัลกอริธึมที่ใช้งานได้ ดังจะเห็นจากทีมงานที่ประกอบด้วยนักวิจัย 4 นาย (Serge Belongie, Chad Carson, Hayit Greenspan และ Malik) ที่พยายามเจริญรอยตามคอนเซ็ปต์ที่เรียกว่า perceptual grouping ด้วยการแบ่งสิ่งที่เป็นภาพออกเป็นส่วนต่าง ๆ (set of regions) ที่มีความคล้ายคลึงกันในเรื่องของสีและ texture โดยให้ชื่อแต่ละส่วนดังกล่าวว่า coherent "blob" ซึ่งแต่ละ blob ประกอบด้วยสิ่งที่แสดง postion, shap, color และ texture blobs ในกรณีนี้จึงเป็นเหมือน representation อันหนึ่งที่ทำหน้าที่เก็บฟีเจอร์ต่าง ๆ ขององค์ประกอบที่เป็นขั้นพื้นฐาน (basic compositional features) สำหรับสิ่งที่เป็นภาพ

คอนเซ็ปต์ในเรื่องของ blob จะเข้ามาทำให้ผู้ใช้ฐานข้อมูลสามารถหาภาพต่าง ๆ ที่มีเครื่องบินชนิดต่าง ๆ บินอยู่ในท้องฟ้า ด้วยการมองหา gray blob ที่อยู่ภายใน blue blob หรือผู้ใช้อาจหาภาพของเสือด้วยการมองหา blob ที่มีสีและ texture ที่เหมือนกับที่ผู้ใช้คาดว่าจะมีในผิวหนังของเสือ รวมทั้ง blob ที่มีสีและ texture ที่เหมือนกับทุ่งหญ้าที่เสือชนิดใดชนิดหนึ่ง

ชอบอาศัยอยู่ ที่น่าสนใจก็คือว่าการนำสิ่งต่าง ๆ มาแสดงให้อยู่ในรูปของ spatial configuration of regions ของสี และ texture เช่นที่กล่าวในข้างต้น สามารถก่อให้เกิดประโยชน์ในแง่ของ machine learnign techniques ก็คือแทนที่จะให้มนุษย์เป็นผู้คิดถึงรายละเอียดของสิ่งที่มีลักษณะเด่นที่สุดของทัศนียภาพที่มีเสือหรือเครื่องบินอยู่ภายในนั้นก็อาจอาศัยวิธีการอื่น เช่น การเรียนรู้ด้วยสถิติ (statistical learning) เข้าช่วย ดังเช่นกรณีของนักวิจัยที่เบิร์กเลย์ซึ่งได้พัฒนาซอฟต์แวร์ที่เรียกว่า statistical classifier ขึ้นมาพร้อมด้วยตัวอย่างของภาพของสิ่งที่มองเห็นได้ในรูปของ categories (เช่นทัศนียภาพที่มีเครื่องบินหรือเสือ) แล้วฝึกให้ซอฟต์แวร์ดังกล่าวสามารถรู้และเข้าใจ (recognize) ภาพที่ไม่เคยเห็นมาก่อนในรูปของปรากฏการณ์ (instances) ของ categories ดังกล่าว ปัจจุบัน classifiers ของเบิร์กเลย์สามารถจำแนกภาพต่าง ๆ โดยอาศัยข้อมูลในด้านของสีและ texture พร้อมด้วยข้อมูลในด้าน region shape เพื่อบอกความแตกต่างระหว่างกลุ่มของ regions ที่มีส่วนในการก่อให้เกิดสรรพสิ่งต่าง ๆ กับกลุ่มของ regions ที่ไม่มีบทบาทอะไร การสามารถเรียนรู้ในกรณีนี้จึงเป็นกลไกทางธรรมชาติที่สามารถเผชิญกับการแปรปรวนต่าง ๆ ที่ไม่เป็นไปตามกฎซึ่งมีอยู่ในระดับชั้นหนึ่งสำหรับสิ่งต่าง ๆ (a class of objcets) การที่ classifier สามารถทำงานโดยปรับตัวเองได้นี้ทำให้อุปกรณ์ทางสถิติแบบนี้ไม่ต้องสนใจกับการแปรปรวนที่ไม่สำคัญ (เช่นสีของท้องฟ้าอาจประกอบด้วย shade ต่าง ๆ ที่เป็นสีน้ำเงินก้อนเมฆอาจมีลักษณะใดก็ได้ ฯลฯ) จึงน่าสรุปได้ว่าเทคโนโลยีที่อยู่เบื้องหลังการเรียนรู้ได้เองนี้จะเป็นอื่นไปไม่ได้นอกจากจะอาศัยสิ่งที่เรียกว่า**นิวรอลเน็ตเวิร์ก** (neural network) ซึ่งได้มีมหาวิทยาลัยอื่น ๆ อีกหลายที่ได้ให้ความสนใจดังเช่นกรณีของ Carnegie Mellon University (CMU) ในสหรัฐฯ ที่ใช้นิวรอลเน็ตเวิร์กในการเรียนรู้และจดจำรายละเอียดต่าง ๆ ที่สำคัญบนใบหน้าคน รวมทั้งโครงการ Photobook ของ MIT ที่จะนำมากล่าวต่อไป

Photobook ของ MIT

นักวิจัยของ CMU ผู้มีนามว่า Takeo Kanade และทีมงานได้ร่วมกันพัฒนาคอมพิวเตอร์โปรแกรม (face recognition software module) ที่สามารถเรียนรู้ลักษณะของใบหน้าของบุคคลต่าง ๆ รวมทั้งการสามารถจดจำลักษณะดังกล่าวที่ถือว่าเป็นอีกตัวอย่างหนึ่งของการใช้ learning techniques ที่จัดว่ามีความสำคัญในแง่ที่ว่า ในขณะที่นักวิจัยรายอื่น ๆ ได้มีความเห็นว่า การสืบเสาะเพื่อทำความเข้าใจในรายละเอียดของใบหน้าของแต่ละคนที่ประกอบด้วย ตา จมูก ปาก หู ฯลฯ เป็นสิ่งที่ทำได้ยากมากนั้น Kanade กับคณะก็สามารถหาทางออกในเรื่องนี้ด้วยการฝึกให้นิวรอลเน็ตเวิร์กสามารถค้นหารายละเอียดทั้งหมดบนใบหน้าโดยไม่ต้องบอกล่วงให้คอมพิวเตอร์รับรู้ว่าโดยตรงว่าบนใบหน้าของมนุษย์ได้ประกอบไปด้วยอวัยวะต่าง ๆ อย่างไรบ้าง วัตถุดิบที่ใช้ในการฝึกนิวรอลเน็ตเวิร์กที่ CMU มีทั้งภาพที่มีใบหน้าคน (คนเดียวและหลายคน) กับภาพที่ไม่มีใบหน้าคน จุดมุ่งหมายของการฝึกก็เพื่อต้องการจะให้เน็ตเวิร์กแบบนี้เรียนรู้สิ่งที่อยู่บนใบหน้าและจดจำเพื่อประโยชน์ในการเปรียบเทียบความคล้ายคลึงของใบหน้าที่ลักษณะต่าง ๆ ในภายหลัง ภาพต่าง ๆ ที่เก็บอยู่ในฐานข้อมูลในกรณีนี้จึงต้องมีมากมาย (large, labeled collection of feces and nonface images) เพราะเป็นโครงการที่เกี่ยวข้องกับเวิลด์ไวด์เวบ รวมทั้งการมีอินเตอร์เฟสที่เปิดโอกาสให้ผู้สนใจสามารถส่งภาพต่าง ๆ มาให้เวบไซท์ทำการตรวจสอบใบหน้า (face detection) ว่าใครเป็นใคร **ความแม่นยำ**ของนิวรอลเน็ตเวิร์กในการแยกแยะระหว่างภาพที่มีใบหน้าคนกับที่ไม่มีใบหน้าคนยังจัดว่าไม่สูงนัก อันหมายถึงว่าจะต้องมีการทดลองต่อไปอีก นอกเหนือจากงานค้นคว้าในด้าน Computerized Facial Recognition (CFR) ที่ CMU ยังมีโครงการอื่น ๆ อีกเช่น การทดลองเพื่อใช้ข้อมูลที่เป็นวิดีโอและออดิโอเพื่อวิเคราะห์พฤติภาพของสื่อที่เป็นวิดีโอและฟิล์ม โดยในกรณีนี้คอมพิวเตอร์จะทำหน้าที่ในด้าน speeds recognition บนออดิโอแทร็ก (audio track) ของข่าวสารโทรทัศน์ เช่น เพื่อให้สามารถบอกชื่อของบุคคลสำคัญที่มีภาพของใบหน้าปรากฏอยู่ในวิดีโออย่างพร้อม ๆ กัน อนึ่ง เรื่องราวของ CFR ในกรณีของ CMU นี้อาจไม่ค่อยชัดเจนเนื่องจากไม่มีการเปิดเผยรายละเอียด อันทำให้ผู้เขียนต้องก้าวต่อไปพูดถึงโครงการอันหนึ่งของ MIT ที่ถือว่าน่าจะประสบความสำเร็จอย่างสูง โดยเฉพาะในเชิงพาณิชย์ นั่นก็คือ Photobook ที่มีความเอกอุในด้าน CFR อันเนื่องมาจากผลของการวิจัยและพัฒนาใหม่ ๆ ที่ทำให้เกิดมี recognition algorithms ที่มีอานุภาพและสามารถใช้กับคอมพิวเตอร์ขนาดเล็กได้

ภายในระบบที่เป็น CFR นั้นคอมพิวเตอร์จะทำหน้าที่ที่สำคัญและสัมพันธ์กันอยู่สามอย่างคือ เวอริฟิเกชัน (verification), เรคกอกนิชัน (recognition) และการหาตำแหน่งของใบหน้าในภาพ (location task) ในขั้นตอนที่เรียกว่าเวอริฟิเกชันนั้นคอมพิวเตอร์จะพยายามแมตช์ภาพสดของใบหน้า (live face) กับข้อมูลที่เป็นภาพที่มีไว้สำหรับการอ้างอิง เรคกอกนิชัน (หรือ identification) เป็นขั้นตอนที่คอมพิวเตอร์แมตช์ภาพสดของใบหน้ากับรูปของใบหน้าต่าง ๆ (saved faces) ที่เก็บไว้ในฐานข้อมูลของคอมพิวเตอร์ สุดท้ายก็เป็น location task ซึ่งคอมพิวเตอร์จะตั้งคำถามว่า "ภาพของใบหน้าอยู่ตรงไหนในภาพ" งานในส่วนนี้จึงสัมพันธ์กับงานที่เป็นเวอริฟิเกชันกับเรคกอกนิชัน เพราะจำเป็นต้องหาตำแหน่งของใบหน้าในภาพดิจิตอลให้ได้ก่อนที่จะเริ่มขั้นตอนในด้านเวอริฟิเกชันหรือเรคกอกนิชัน เป้าหมายของผลิตภัณฑ์ เช่น Photobook หรือระบบ CFR อื่น ๆ จึงไม่ได้อยู่ที่การทำหน้าที่ดังกล่าวแต่อย่างเดียว แต่จะต้องสามารถดำเนินการให้ได้ในรูปของรีลไทม์ (real-time) หรือใกล้เคียงกับรีลไทม์ Photobook สามารถวิ่งได้ในยูนิกซ์แพล็ตฟอร์ม นอกจากนี้ก็มีผลิตภัณฑ์อื่นเช่น Sherlock ซึ่งเป็นเรคกอกนิชันซอฟต์แวร์ที่ให้การสนับสนุน DOS, Windows และ OS/2 แพล็ตฟอร์ม

การทำงานของ Photobook ได้อาศัยความรู้ขั้นพื้นฐานจาก information theory ในแง่ที่ว่าในขั้นแรกโปรแกรมตัวนี้จะแบ่งแต่ละใบหน้าของคนออกเป็นส่วนต่าง ๆ ที่เป็นสองมิติ ซึ่งประกอบด้วยส่วนที่สว่างและมืด จากนั้นอัลกอริธึมภายในโปรแกรมก็จะหาฟีเจอร์อันเป็นจุดต่าง ๆ ที่ถือว่าเด่นที่สุดบนใบหน้า เพื่อหาความแตกต่างของใบหน้าหนึ่งกับใบหน้าอื่น ๆ ฟีเจอร์ที่เป็นจุดเด่นนี้มีชื่อเรียกว่า ไอเก้นเฟส (eigenfaces) จากนี้อัลกอริธึมก็จะแทนใบหน้าแต่ละใบหน้าด้วยส่วนประกอบที่เป็นไอเก้นเฟสต่าง ๆ ซึ่ง Photobook จะเก็บสิ่งที่เป็นไอเก้นเฟสของแต่ละใบหน้าไว้ในฐานข้อมูล ในการใช้งาน เมื่อมีใบหน้าของบุคคลใดบุคคลหนึ่งส่งเข้ามาให้ตรวจสอบ (target face) โปรแกรมก็จะเปรียบเทียบคุณสมบัติของไอเก้นเฟสของใบหน้าใหม่นี้กับไอเก้นเฟสในฐานข้อมูล โดยอัลกอริธึมจะเลือกใบหน้าต่าง ๆ ที่เห็นว่ามีลักษณะใกล้เคียงกับ target face และสมมติว่าได้มีการกำหนดมาตรการในการยอมรับความเหมือนของใบหน้าที่ส่งเข้ามาตรวจสอบกับใบหน้าที่แท้จริงของบุคคลใดบุคคลหนึ่ง (recognition thresh old) ที่มีได้เพียงใบหน้าเดียว รวมทั้งการเปรียบเทียบในแง่ของแมตชิงซึ่งสามารถผ่านข้อต้องการของ threshold ดังกล่าวได้อย่างครบถ้วน สิ่งที่เป็น target face ก็จะกลายเป็นภาพของใบหน้าที่คอมพิวเตอร์จำได้ว่ามีอยู่จริงในฐานข้อมูล ซึ่งไม่ใช่ใบหน้าแปลกปลอม ในทางกลับกันโปรแกรมตัวนี้ (Photobook) ยังสามารถดิสเพลย์ **matching face** ต่าง ๆ ให้ผู้ใช้ได้แลเห็นด้วยการเรียงลำดับของการแมตช์ เพื่อให้ผู้ใช้เป็นผู้คัดเลือกสิ่งที่เป็น target face เอาเอง ในแง่ของความแม่นยำนั้นได้ปรากฏในการทดลองครั้งหนึ่งว่า ในฐานข้อมูลที่มีภาพใบหน้าอยู่ 7,562 ใบหน้านั้น โปรแกรมสามารถให้ recognition rate ที่สูงถึงร้อยละ 95 สำหรับในการใช้เนื้อที่เพื่อเก็บข้อมูล ถ้าใช้ floating-point number ขนาด 2 ไบต์ ให้ทำหน้าที่แทนแต่ละค่าของไอเก้นเฟส ขนาดของข้อมูลของแต่ละใบหน้าก็จะเท่ากับ 80 ไบต์ เมื่อเทียบกับการไม่ใช้ไอเก้นเฟส ซึ่งข้อมูลจะมีขนาดใหญ่ถึง 250 KB (ก่อน compression) สำหรับแต่ละใบหน้า ความรวดเร็วในการเปรียบเทียบหรือตรวจสอบแต่ละใบหน้าในกรณีของ Photobook นี้จะใช้เวลาราวหนึ่งวินาที นอกจากนี้ก็ยังมีผลิตภัณฑ์อื่น เช่น TrueFace ที่เป็นระบบ CFR ซึ่งใช้นิวรอลเน็ตเวิร์กเทคโนโลยีโดยใช้กับพีซีบนฐานเพนเทียม ให้เลือกใช้ในราคาราวชุดละ 5,000 เหรียญสหรัฐฯ

สรุป

พัฒนาการและการแก้ปัญหาต่าง ๆ ที่เบิร์กเลย์ในการค้นคว้าวิจัยเพื่อค้นหาภาพที่เป็นดิจิตอล จากฐานของเก็บที่มีขนาดใหญ่มาก ๆ (large colections) ซึ่งได้เน้นในบทความนี้ ยังมีที่ใช้ในด้านอื่น ๆ อีก เช่น การกลั่นกรองข้อมูลข่าวสารที่ได้มาจากเวบ รวมทั้งการดิจิไตส์สิ่งที่เป็นภาพและฟิล์มในรูปของ large collections ซึ่งกำลังเป็นที่นิยมในขณะนี้และในต่อไป ในแง่ของการเป็นเนื้อหา (content) ให้แก่สื่อที่มีการเปลี่ยนแปลงเร็ว เช่น ข้อความสำหรับอีเมล์และเวบเพจต่าง ๆ สิ่งที่ทีมงานที่เบิร์กเลย์ต้องการจะปรับปรุงให้ดีขึ้นสำหรับในการค้นหาข้อมูลที่เป็นภาพก็คือ ความรวดเร็วในการให้ข้อมูล (retrieval rate) จากฐานข้อมูลเพื่อให้โครงการเช่น Digital Library Project ทำหน้าที่เป็นห้องสมุดที่มีประโยชน์และมีประสิทธิภาพสำหรับผู้ใช้ จนในที่สุดการวิจัยในเรื่องของการค้นหาภาพที่เป็นสามมิติ ซึ่งทีมงานของเบิร์กเลย์กำลังดำเนินการอยู่นั้นจะก้าวไปถึงจุดที่ว่า ผู้ใช้ฐานข้อมูลสามารถที่จะค้นหาภาพเช่นของเสือดาวที่กำลังกระโจนวิ่งอยู่ในทุ่งโล่งได้ในเวลาอันสั้น ด้วยการออกคำสั่งเพียงไม่กี่คำ (เช่น "อยู่ในตระกูลแมว "มีแต้มจุดสีดำบนหนังสีเหลือง" "กระโจนวิ่งได้เร็ว") อันเป็นหนึ่งในปัญหาที่ลุ่มลึกสำหรับในเรื่องของคอมพิวเตอร์หรือแมชีนวิชัน

เส้นทางในการค้นคว้าวิจัยในด้านแมชีนวิชันน่าจะแบ่งออกได้เป็นสองส่วนคือ pure vision research ที่ไม่อาศัยความรู้ทางชีวภาพเข้าช่วย กับการวิจัยในอีกส่วนหนึ่งที่พยายามจะอธิบายการทำงานของระบบวิชันของสรรพสัตว์ อย่างไรก็ตาม งานวิจัยในด้านแมชีนวิชันซึ่งถือว่าเป็นสาขาหนึ่งของ AI ยังต้องเกี่ยวข้องกับธรรมชาติเป็นอย่างมาก โดยเฉพาะในเรื่องของสิ่งมีชีวิตที่อยู่ในรูปของวิทยาศาสตร์ชีวภาพ ดังจะเห็นได้จากการค้นคว้าในด้านวิชันในอดีตซึ่งเป็นที่มาของคำว่า **"hierarchy of representations"** ซึ่งในตอนแรกก็ไม่ได้มีการให้ความสำคัญแก่ "คำ" ดังกล่าวว่าได้มีส่วนเกี่ยวข้องกับงานวิจัยในด้านสัตว์ศาสตร์ จนกระทั่งในเวลาต่อมาจึงเริ่มยอมรับกันว่าภายในสมองของสัตว์ (animal brains) ได้มีหลักขั้นพื้นฐานในการทำงานที่อยู่ในรูปของ hierarchical representations เหมือนระบบแมชีนที่มนุษย์สร้างขึ้นมา นอกจากนี้อิทธิพลของระบบทางชีวภาพก็ยังมีต่อไปอีกในเรื่อง parallel computation ที่มีบทบาทอย่างมากในเรื่องของการวิจัยอย่างเป็นงานเป็นการในด้านวิชัน เมื่อเป็นเช่นนี้งานค้นคว้าในด้านแมชีนวิชันที่กว้างไกลจึงควรอยู่ในรูปของ cross-disciplinary research efforts เพื่อให้สามารถสร้างระบบวิชันที่เหมาะในทางปฏิบัติและมีความรวดเร็วในรูปของรีลไทม์ (pratical, real-time visual perception)